# Optimization of LTM Networks Using GTS: Statistical Approaches to Spatial and Temporal Redundancy

Kirk Cameron, Ph.D.
MacStat Consulting, Ltd.

Philip Hunter, P.G.
Air Force Center for Environmental Excellence (AFCEE)

## Background

The Air Force, in an effort to wisely program and more cost-effectively manage resources for long-term monitoring (LTM), has developed a spatial and temporal optimization algorithm based on statistical methods. A major requirement was that the optimization algorithm should be reasonably simple to implement, yet effective in identifying temporal and spatial redundancy. The algorithm also had to be structured in terms of a decision-logic flowchart that would be useful more generally at additional sites across the Air Force and the Department of Defense. Components of the algorithm are also useful for monitoring long-term operations of remedial systems that may currently be in place.

Data from two ground-water plumes at the Massachusetts Military Reservation (MMR) on Cape Cod were used to test and develop the statistical approaches. Monitoring networks associated with the FS-12 and Eastern Briarwood Plumes were analyzed. The database consisted of 173 distinct well locations for FS-12 and 363 distinct well locations for Eastern Briarwood. Both of these plumes were sampled over a period of years. Contaminants analyzed include ethylene dibromide (EDB), benzene, trichloroethylene (TCE), and tetrachloroethylene (PCE). Other contaminants either rarely or never exceeded applicable drinking water standards (MCLs) and so were excluded from the analysis. As is the case for most Air Force installations, almost 90% of the measurements were non-detect.

The key question of interest was: how can LTM networks be optimized so that limited resources are wisely and effectively dedicated to field sampling, laboratory analysis, and/or well construction? The primary objective was to determine to what degree these resources can be pared without significant lose of key statistical information about the plumes being monitored.

In particular, it was assumed that the goal of any LTM effort is to provide an accurate assessment over time of ground-water quality, with the ultimate objectives of enabling one to 1) construct an interpolated map of the current concentration levels across the site area, and 2) accurately assess trends or other changes in individual monitoring wells. Interpolated maps are used to assess whether or not a plume of contaminated ground water exists, and, if so, its extent and characteristics (e.g., intensity). Changes in such maps over time can indicate either improvement or decline in ground-water quality across the plume area. Changes in concentration patterns or the identification of trends at individual "sentinel" wells can also serve the same purpose.

The optimization algorithm itself was divided into two separate components: 1) temporal redundancy, and 2) spatial redundancy. Temporal redundancy refers to whether or not certain wells are being sampled too frequently. Are samples collected so often that there is a significant degree of autocorrelation between closely spaced measurements? If so, can this redundancy be reduced or eliminated by decreasing the frequency of sampling and/or lengthening the time between collection of samples? Spatial redundancy refers to whether or not too many wells are being monitored. That is, are there wells that provide essentially redundant information and could be eliminated from the network without sacrificing resolution of ground-water quality?

At the root level, the optimization algorithm presented below consists of three basic steps: 1) Identification of temporal redundancies in currently monitored wells; 2) identification of spatially redundant wells; and 3) projection of cost savings gained by eliminating wells and/or reducing sampling frequencies.

## Temporal Optimization Algorithm

The temporal algorithm was divided into two non-overlapping pieces: A) computation of the composite temporal variogram, and B) "iterative thinning" of sampling events at selected wells.

The first approach allows time series data from many wells to be combined together into a single measure of temporal autocorrelation known as a variogram. As opposed to spatial autocorrelation, which considers the distance between points in space, temporal autocorrelation takes "distance" as the elapsed time between samples collected from the same location. Using time as the distance function, one-dimensional variograms can be constructed to measure the average correlation between pairs of measurements as the time lag between them increases or decreases.

The point in time at which the "sill" (i.e., upper bound) of this variogram is reached estimates the approximate lag time between sampling events for which there is no time-related dependency, and hence, no temporal redundancy. Samples taken from the same location at shorter intervals will tend to be correlated to some degree and therefore at least partially redundant in the statistical information they provide. Because data from multiple well locations are included in the composite temporal variogram, the variogram approach does not necessarily provide an optimal sampling frequency for each individual well. Rather, it estimates an "average" optimal frequency that can, if desired, be adopted on a site-wide basis.

To estimate the average temporal autocorrelation, the basic approach of Tuckfield (1994) was modified to estimate a one-dimensional variogram using time of sampling as the dimension. Instead of trying to explicitly model the temporal autocorrelation, the key steps were to 1) compute an empirical temporal variogram for each well; 2) average the empirical variograms across wells to build a composite temporal variogram; 3) locate the smallest time interval at which the approximate sill of the composite variogram was reached; 4) designate the time interval found in Step 3 as the minimum sampling interval providing essentially uncorrelated temporal data; and 5) adjust the sampling frequencies at the remaining monitoring wells so that the time lag between samples will not fall below this minimum interval.

The second approach is designed for key "sentinel" wells, wells exhibiting trends over time, or other monitoring locations for which a well-specific sampling frequency may be desirable or needed. Various methods for optimizing sampling frequencies at individual wells have been proposed (see Johnson, et al, 1996 for instance). With this algorithm, however, a somewhat different approach was taken. "Iterative thinning" refers to the temporary removal of randomly-selected data points from the time series of measurements at a given well. The algorithm consists of 1) estimating a trend using the entire time series, 2) thinning the time series by a fraction of the measurements, and then 3) re-estimating the trend to determine if the slope estimate is still close to the original slope. Additional thinning can occur until the "thinned" trend estimate is significantly different from the original trend.

Problems in the analysis were encountered for a small subset of the wells which did not exhibit a readily discernible trend or lacked enough time series data to make a well-specific analysis worthwhile. Further complicating the matter, at some of these wells the apparent trends were non-linear, or, even when the trend was fairly linear, in some cases the concentration data were much better behaved (i.e., exhibited less variation around the trend line) toward the end of the sampling record than the near the beginning.

For all these reasons, it is recommended that the iterative thinning approach be restricted to selected wells at a given site having adequate data and some indication of a trend if possible. Furthermore, to avoid statistical assumptions inherent in standard linear regression methods, trend estimation was done with a non-parametric technique known as Sen's method (Gilbert, 1986). Sen's procedure can be applied to a wide variety of datasets and is readily adapted to non-detect measurements and irregular sampling frequencies.

Using Sen's method, only the basic linear slope of the time series was estimated, along with a confidence interval around the slope estimate. The premise of the iterative thinning approach is that if fractions of the data are randomly removed from the time series, yet the same basic slope is estimated (within the bounds of the original confidence interval) on the reduced dataset, temporally redundant data exists and the sampling frequency at that well can be adjusted to further lengthen the time between sampling events.

**Steps in Temporal Variogram Approach**

**1) Pre-Process the Data**

Several steps were necessary to prepare the datasets for optimization including filling data gaps in the time series data and populating various missing data fields. One cannot overestimate the importance of database management and the use of a standardized electronic data structure in attempting this kind of analysis. Indeed, on the order of 80% of the effort involved merely producing a "clean" dataset that was worthy of applying the statistical methods for the analysis. Data was produced by disparate sources including several consulting firms and government agencies. An electronic data structure modified from the Air Force's Environmental Resources Program Information Management System (ERPIMS) was used to create the data warehouse. Inconsistencies in laboratory qualifiers, various detection limits, differing analytical methods, non-uniqueness in well identifiers were all challenges to developing a dataset for analysis.

Since the analytical methods and detection limits varied somewhat from sample to sample and across the years of data collection, and the vast majority of the data were non-detect, it was impractical to fit standard parametric distributional models to the concentration data. Instead the data were simplified by transforming each reported value into an "indicator value" (IV), that is, a zero or one respectively, depending on whether the value exceeded a fixed concentration cutoff for the contaminant.

With an indicator transformation, information about extreme concentration levels is lost (other than knowing the value exceeds the cutoff). However, it is often easier using indicators to fit the kinds of geostatistical covariance models (discussed later) needed to gauge the degree of spatial redundancy and to determine an approximate "sill" for measuring the lowest point of temporal redundancy. Furthermore, non-detect concentrations need not be known or imputed (at least if the detection/quantitation limit is at or below the cutoff), since any concentration presumably less than the detection/quantitation limit also will not exceed the chosen cutoff. It is therefore possible to unambiguously classify a dataset into indicator values without resorting to complicated imputation schemes or tenuous statistical models.

While the concentration cutoff used to form the indicator values is somewhat arbitrary, "natural" options would include the highest detection or quantitation limit, an applicable MCL or regulatory limit, or perhaps an already established background level (e.g., a mean or upper confidence limit). At FS-12, the indicator cutoff was taken as the MCL: 5 ppb for benzene and .02 ppb for EDB. At Eastern Briarwood, the cutoffs for TCE and PCE were selected equal to their respective detection limits. For the latter plume, a sill was more readily identified for both contaminants using the detection limit as the cutoff instead of the MCL of 5 ppb. In practice, more than one cutoff may need to be examined to ensure that the least amount of statistical information is lost when forming either the temporal variogram or the spatial variograms to be discussed later.

**2) Compute Composite Temporal Variogram**

Once the data were converted to indicators, a sample estimate of the time-dependency between sampling events was computed known as the composite temporal variogram. A temporal variogram is a measure of correlation over time between two sampling events (at the same well), roughly equal to the average squared difference in indicator values for all pairs of measurements separated by a given "lag" (i.e., defined here as the time between sampling events).

To form the composite temporal variogram, separate variograms were first calculated using the time series from each individual well. First, a base lag spacing was chosen to represent increasing periods of time. For example, one set of lags might be taken as 0 months (1$^{st}$ lag), 2 months (2$^{nd}$ lag), 4 months (3$^{rd}$ lag), 6 months (4$^{th}$ lag), and so on, using a lag spacing of 2 months. Then, at each distinct well location, the squared differences between indicators from all possible pairs of sampling events were computed and grouped by nearest lag. After averaging the squared differences associated with each distinct lag, a sample variogram for each well was born. Finally, a composite temporal variogram was computed by averaging the individual well variograms across wells for each common lag to get a typical measure of temporal correlation applicable to the site as a whole.

One complication encountered in constructing the composite temporal variogram was that the historical records at many wells were quite limited, with less than a handful of separate sampling events at many locations and very tight temporal spacing at others (e.g., all samples collected over a two-to-three week period). Because of this, many wells contributed only a small number of data pairs to the composite variogram, at perhaps one or two lags. Fortunately, other wells contributed longer data records, allowing the composite variogram to be "filled out" with additional time lags.

Since the actual times between sampling event pairs often did not correspond exactly to a given set of lags, the same calculations explained above were made for three different base lag spacings. Then the variograms from all three spacings were amalgamated together to get the final composite variogram. This tends to ensure that the resulting temporal variogram is not biased due to an artifact of choosing one particular set of lags.

Another point to note is that for nested wells (i.e., wells with multiple screens at different depths but along the same bore hole), concentration values from different depths were converted to separate indicator values and used independently in forming the temporal variograms. That is, the well data were not stratified by depth. One consequence of this simplifying step was that the variation measured by the temporal variogram could be somewhat overestimated, especially at the smallest lags. However, the degree of overestimation should be partially offset by the use of indicator values instead of the actual concentrations in the variogram computations.

**3) Adjust Global Sampling Frequency**

With a composite temporal variogram in hand for each contaminant of concern, a non-linear smoothing technique was applied to each graph to determine an approximate sill. The sill represents the highest *stable* numerical level on a variogram. It first occurs at the smallest lag time where there is no discernible correlation between a pair of sampling events. That is, if a sill has been reached when reading a variogram from left to right, any pair of sampling events that are separated by lag times at least as long as those associated with the sill should be uncorrelated in a statistical sense.

Given the fact that some information is lost by converting the actual concentration data to indicators, it might be tempting to use the actual concentrations when forming the temporal variogram. However, it must be remembered that the composite variogram is an average of the temporal variation from all wells at the site. The same precise temporal pattern is not likely to hold for each and every well. Indeed, the composite variogram is designed to be a parsimonious way to determine a typical or "average" global sampling frequency that can be applied more or less to all the wells uniformly. As a consequence, though, one should not expect to see simple, smooth patterns when examining the points on a composite temporal variogram.

Using the indicator values to form the temporal variograms resulted in substantial variation in the estimated values at neighboring lags. The same variograms computed on the raw concentration data resulted in even greater variation, so this avenue was not pursued further. What was done, however, was to apply a non-linear smoothing algorithm to each composite variogram, in order to estimate a smooth pattern consistent with the data. A variety of non-linear smoothers are available in standard statistical software packages, including moving window averages, geostatistical variogram fitters, the

Levenberg-Marquardt algorithm, etc. In this case, a lowess procedure (lowess denoting a **lo**cally-**we**ighted regre**ss**ion) was applied to the variograms.

The lowess procedure is akin to a moving window average, but instead of a simple arithmetic average a weighted regression is performed on the data points included in each moving window. Like all moving window algorithms, the resulting smooth depends on the size of the window used, so alternate window widths must typically be tried to properly balance the degree of smoothness in the fit and how quickly the fit responds to changes in the data.

The purpose for using any smoother on the composite temporal variograms is to try and identify an approximate sill, and to determine at what approximate lag time the sill first occurs. The ability to identify the sill varied with the respective contaminants. The lowess fit approach showed that concentrations first leveled off in the approximate range of 350-500 days, depending on the the analyte. Some difficulty in ascertaining a sill was due to the influence of the indicator cutoffs assigned to the various contaminants. Recall that indicator cutoffs were either set at detection limits or at MCL levels. The cutoffs that were set to the highest detection limits resulted in the more readily discernible sills for those particular contaminants involved.

Remembering that the sill on a temporal variogram represents the point of lowest correlation between sampling events (actually the point of zero correlation), the smallest lag time associated with the sill for a given contaminant can be taken as a kind of optimal sampling interval, optimal in the sense of indicating the shortest time between samples with zero statistical correlation. Any shorter interval is associated with some temporal redundancy, since the correlation for such lag times is positive. Consequently, the results of this approach suggest that the general lag time between samples at FS-12 should be at least a year (approximating to the nearest quarter for operational simplicity), while that for Eastern Briarwood should be at least five quarters or 1.25 years.

The application of the proposed optimized schedule to the current sampling frequencies at Eastern Briarwood — without removing any of the current monitoring wells — would result in a 36% annual reduction in the total sampling and analysis budget. Even greater cost reductions are projected for the FS-12 plume; however, the savings are based on reduction in sampling frequencies as well as removal of certain currently monitored wells.

Note that this approach does not adjust the sampling frequencies of individual wells. Rather, the composite temporal variogram offers a “broad brush” view of temporal autocorrelation on-site and provides an impartial method to set uniform, optimal sampling frequencies based on minimizing the degree of temporal autocorrelation. Of course, selected wells may need to sampled more often for other reasons (e.g., new well installations, hydrogeologic factors, etc.). And there may be some wells with well-defined trends that can be adjusted/optimized individually using the other approaches above. However, in the test case applied to these particular plumes, if there was difficulty in locating a discernible trend in the contaminants of concern, the composite temporal variogram holds the promise of estimating a typical (albeit rough) temporal pattern that can facilitate sampling decisions.

**Steps in Iterative Thinning Approach**

**1) Establish Baseline Trend**

To establish an initial trend estimate for a given well, the slope was estimated using a non-parametric technique known as Sen's method. Sen's slope estimate involves picking the median of a list of pairwise slope values, where each possible pair of data values is used to calculate a pairwise slope.

To account for sampling and measurement fluctuations/variability, basic formulas involving the same list of pairwise slope values were used to compute a confidence interval around Sen's slope estimate (Gilbert, 1987). Then, by constructing a confidence interval around the initial trend, re-estimates of the trend after "thinning" the data series could be compared to the confidence bounds to determine whether the slope value had changed in a significant way.

As noted earlier, Sen's method can be adapted to the presence of non-detects, although some choice must be made to impute the non-detect concentration values. Perhaps the easiest tack is set all non-detects equal to zero for purposes of estimating the trend. Other values might be chosen, such as half the detection or quantitation limit. However, if multiple detection/quantitation limits exist in the data, one should be careful not to estimate a positive or negative pairwise slope between two non-detects just because their detection limits are different. In this study, all non-detects were treated as zeros, so that any pairwise slope calculated between two non-detects was necessarily zero as well.

**2) "Thin" the Data Series and Assess Accuracy**

Once Sen's slope estimate and the confidence interval around the trend were in hand, the data series was "thinned." To do this, a column of random numbers between 0 and 1 was generated alongside the time-ordered concentration data. Then, in iterative fashion, increasing percentages of the data were randomly "removed" from each time series. For example, at 20% censoring, from each successive group of five measurements one was removed, simply by flagging the lowest random number from the corresponding column in that group of five. At 33% censoring, one of every three successive values was removed, and so on. Flagging the values in this way ensured that the random removals would not be "bunched" at one end or the other of the time series; rather, the series was simply "thinned out" in a practical way.

After thinning each time series, Sen's slope (but not the confidence interval) was recomputed to see if it fell within the original confidence bounds and to make sure the sign of the slope had not changed. The highest censoring level for which the re-computed slope was still comparable to the original trend estimate was then used to adjust the sampling frequency at that well and to determine the degree of temporal redundancy that existed.

**3) Adjust the Well-Specific Sampling Frequency**

To optimize the sampling frequency at a given well, the fraction of data points removed in the thinning process was considered. For instance, if the highest level of thinning was 50% before the slope changed, half the data could be removed and yet still provide a comparable slope estimate. In this case, the optimized sampling interval would essentially double in length. If only 20% of the data were removable, the optimized sampling interval would increase by roughly 25%.

More fundamentally, the optimized sampling interval for wells sampled on a fairly regular schedule can be computed as the total length time in the sampling record divided by the number of points remaining after thinning. For wells with irregular sampling histories, it is important to avoid biasing the sampling frequency by early periods of intense sampling. Consequently, the optimized sampling interval should be computed by dividing the most recent sampling interval (i.e., the lag between the two most recent, distinct sampling dates) by one minus the fraction of data points thinned. This step should avoid the problem of creating an optimized sampling interval that might actually be *shorter* than the current one based on the most recent monitoring schedule. However, some caution must be used when performing such a step on an automated basis. Ideally, the interval from the most current sampling schedule ought to be utilized before dividing by the complement of the fraction thinned, rather than just assuming that the last two sampling events adequately define the sampling interval, as in this report.

Results of this pilot study at FS-12 and Eastern Briarwood suggest that temporal redundancies did indeed exist, at least for selected wells with sufficient data. Data at these wells could be "thinned" without losing the ability to estimate the basic trend in concentration levels over time. The iterative thinning approach is fairly easy to implement and does not require more sophisticated non-linear fitting of the trend function. However, it does presuppose a sufficient number of data values (say at least 8 to 10) with which to perform the random subsetting and to estimate the slope of the trend. The average fractions thinned varied in the range of 40 to 70 percent. Such thinning percentages can potentially translate into significant cost savings in reduced sampling at the impacted wells.

## Spatial Optimization Algorithm

The spatial side of the optimization algorithm is predicated on the notion that well locations are redundant if nearby wells offer nearly the same information about the underlying plume. Specifically, a well is considered redundant if its removal does not significantly change an interpolated map of the plume; that is, essentially the same iso-concentration contours result.

The path taken in identifying potentially redundant wells included the following steps: 1) generate an initial plume map via a geostatistical interpolation method known as kriging; 2) assign numerical weights (denoted *global kriging weights*) to the well locations in the monitoring network to gauge their relative contribution to the plume map; 3) temporarily remove that subset of wells with the lowest global kriging weights and re-estimate the plume map; and 4) assess whether the plume map has changed in any significant way and gauge via the kriging variance whether the spatial uncertainty has substantially increased. If not, try removing some additional wells and repeating the process. But if significant changes are evident, do not remove that subset of well locations.

### Key Steps

#### 1) Pre-Process the Data

Many of the same initial steps used to form the composite temporal variogram were taken to prepare the data for spatial optimization. The data at both plumes were also converted from

concentration values into indicators (IV), using the same cutoffs as selected for the temporal optimization.

To keep the statistical algorithm as operationally feasible as possible, only a single indicator cutoff (e.g., MCL, detection/quantitation limit, etc.) was used to convert the raw data. One could potentially use a multiple indicator approach with multiple cutoffs set at increasing concentration levels. Less detail about the extreme portions of the plume would be lost. However, the steps needed to geostatistically model the data would be multiplied. Unless detailed information about the plume is needed, the added complexity will probably not provide much in the way of useful information on spatial redundancy over the approach taken here.

Another simplifying step was applied to nested wells that were screened across multiple depths. Since the number of data points at any given depth was fairly limited, the three-dimensional nature of each plume was collapsed into a two-dimensional problem by "averaging" indicator values over depth for a given well location. "Averaging" in this context refers to labeling the sampling at a particular well on any given date/event as a "hit" (i.e., exceeding the indicator cutoff, so that IV = 0) if any one or more of the samples with depth was a "hit." Well locations where all values did not exceed the indicator cutoff were assigned indicator values of IV = 1.

A final pre-processing step was necessary to accommodate the irregular sampling schedules observed at the plumes. Some of the wells were sampled more often than others. To avoid giving more statistical weight to some well locations than others simply by the volume of data points available at frequently sampled wells, each dataset was divided arbitrarily into a series of quarterly "snapshots" or time slices.

For a given three-month time span, a well with *any* "hits" was labeled as a "hit" for that quarter, regardless of the number of times it was sampled. This meant that as long as a well was sampled even once during that quarter, it was given the same relative statistical weight as a well sampled more frequently. Again, collapsing the problem in this way "loses" or ignores some information about the temporal pattern of contaminant concentrations. But the gain in simplicity is significant. A related decision was made to limit the number of time slices included in the spatial analysis to those with a relatively large number of wells sampled (typically 30 or more).

**2) Model the Spatial Covariance**

Once the data were collapsed into a single horizontal plane and grouped by quarter of sample collection, the indicator data from each quarterly time slice were fit to standard geostatistical spatial covariance models. This involved two basic steps. First, a sample estimate of the spatial correlation function known as the empirical variogram was computed. A variogram is a measure of correlation with distance between two sampling locations, roughly equal to the average squared difference in indicator values for all pairs of locations separated by a given "lag" (i.e., distance between locations).

To account for possible changes in the plumes over time, the empirical variogram for each quarterly slice of data was examined to see if that pattern also changed with time. Because the correlation pattern was fairly similar for the bulk of the quarterly time slices, the differences were not great enough to

necessitate separate variance modeling. Consequently, the quarterly variograms were averaged across the time slices (weighted by the number of pairs contributed to the variogram estimate in each time slice) to form a single, time-averaged variogram.

Another typical step is to compute the empirical variograms with different base lag spacings (e.g., separations between successive lags of 100 ft, 300 ft, 1000 ft, etc) to ensure that the choice of lag does not overly influence the appearance of the resulting variogram. In this study, the basic pattern was similar regardless of the base lag spacing employed. So, the empirical variograms at three different lag spacings were amalgamated to form a final variogram prior to modeling.

Once the empirical variogram for each contaminant was estimated, a non-linear fitting program (utilizing the Levenberg-Marquardt algorithm) was used to determine an appropriate positive-definite spatial covariance model. Four such models are common in the geostatistical literature: spherical, exponential, gaussian, and power. The fitting algorithm was set up to either fit a combination of up to three spherical, exponential, and/or gaussian components (termed "nested structures" in geostatistical parlance) or a combination of up to three power model components.

At FS-12, EDB was best fit with essentially a spherical model, while the variogram for benzene was best described by a power model (with a nearly quadratic power coefficient). At Eastern Briarwood, TCE was best fit with a power model (using a nearly quadratic power coefficient), while for PCE, no standard model was entirely adequate. Instead, the variation with distance for PCE fluctuated more or less around a constant level, suggesting that a constant "nugget" variance term be used as the variogram model. Such a model implies that there is no correlation with distance between neighboring well locations. Lack of spatial correlation might be related in this setting to the indicator cutoff used for PCE and/or the "averaging" of the indicators across depths within wells or across sampling events within time slices. In any event, the results of spatial redundancy analysis at Eastern Briarwood when comparing TCE and PCE, as discussed below, were surprisingly consistent despite the difference in spatial model used.

**3) Krige the Indicator Data**

To actually determine which wells might be spatially redundant, remember that a monitoring well was considered redundant if other (nearby) wells provided much the same information concerning the plume being monitored. More specifically, a well was redundant if it provided little independent or additional information when generating a map of the plume.

To generate a plume map, estimates of concentration are needed at unsampled locations, not just the installed wells. These estimates typically involve an interpolation of the known concentrations at already existing wells. Most often the interpolation is computed as a weighted linear combination of the sample data from a series of ($n$) fixed locations.

One way to define spatial redundancy is when one or more wells are assigned very small weights in the interpolation process compared to other wells. For only those wells with larger weights contribute significantly to the estimation at unsampled locations.

While many methods for linear interpolation exist, the approach adopted at MMR is a widely used method for linear interpolation over a spatial area known in the geostatistical literature as *kriging*. Not only does kriging offer a kind of "best" unbiased linear interpolation, but it alone among common interpolation schemes explicitly accounts for the statistical redundancy at nearby sample locations, through its estimate (i.e., model) of the spatial covariance function.

Although there exist a variety of kriging methods, the most common is known as *ordinary kriging* (OK). OK generally works well for concentrations or other measurement data that are mostly detected or quantified. For chemical parameters with low detection rates, however, OK can be difficult to apply for three reasons: 1) the spatial covariance model is often hard to model due to the unknown concentrations of non-detects and the fact that some type of imputation must be made for these unknown values; 2) the linear interpolation at unsampled locations must again rely on combinations of imputed non-detect values; 3) if the detected values are additionally quite skewed, both the empirical variograms and kriged estimates can fluctuate in unpredictable and anomalous ways.

Empirical variograms of the raw concentration data (simply taking non-detects as zeros) were quite jagged and impossible to fit using the standard spatial covariance models. Because of this and the reasons outlined above, an alternative procedure known as *indicator kriging* (IK) was employed. In simple IK, all the sample data are re-classified as ones or zeros depending on whether or not the actual concentration is below or above a fixed threshold (i.e., the indicator cutoff). While detail concerning the intensity of the plume is lost, there is also no need to know the exact concentrations of non-detects. Furthermore, the algorithm is exactly the same as OK except that the indicator data are used in place of the raw concentrations.

Because the data are transformed into indicator values, the results of IK interpolation at unsampled locations do not represent concentration estimates. What they do represent are probabilities of exceeding or not exceeding the indicator cutoff. That is, assuming for instance that the MCL is the cutoff, a *low* IK estimate denotes a *low probability* that the true concentration at that estimated location is *below* the MCL, while a *high* IK estimate denotes a *high probability* that the true concentration *is below* the MCL. Low IK values therefore represent probable MCL exceedances, while high IK values represent the opposite.

The basic IK algorithm used at MMR was to first divide the plume area into a series of non-overlapping blocks. At each block a simple search algorithm was used to locate a set of sampled locations closest to the block. Then, using the modeled spatial covariance function, local kriging or interpolation weights were computed based on the spatial configuration of the known indicator values (that is, the data from known surrounding wells) relative to the block and the spatial correlation between the average block location and each known indicator. These local weights were then combined with the known indicator values to generate a block indicator estimate (consisting of a weighted average of the indicators). The block indicator estimates taken as a whole produce an estimated indicator plume map.

**4) Compute Global Kriging Weights**

Though the indicator plume maps resulting from indicator kriging do not provide explicit concentration estimates for the plume, two intermediary computations from the kriging exercise are

extremely useful: 1) the local kriging weights assigned to sampled locations near each block can be accumulated and averaged to generate a "global" interpolation weight for each well (see Isaaks & Srivastava, 1989); 2) at each block, the local kriging estimation variance indicates the relative uncertainty of the local block estimate compared to estimates at other blocks.

Since the search algorithm described above when kriging individual blocks locates known wells nearest to the block being estimated, some wells, depending on their location (and particularly those toward the middle of the plume), are used in the local estimation of many different blocks. These wells will potentially receive a different *local* kriging weight each time they are tagged by the search algorithm, since the geometric position of a fixed well relative to the block being estimated will change with each new block. The global interpolation weights were thus formed by averaging all of the local kriging weights for each given well, in order to estimate the well's overall contribution to the estimation process. Note that to assess the average contribution of a given well not only across the site spatially, but also over time, the local kriging weights for each time slice were further averaged across the slices.

The global interpolation weights offer a relative ranking of the well locations in terms of the amount of independent spatial information provided. Those wells that are spatially redundant will tend to have the lowest global weights since their local kriging weights will frequently be small. By choosing a threshold value of, for instance, .01 or .02, and eliminating all those wells with global weights no greater than the threshold, an impartial decision criterion can be established for removing spatially redundant well locations.

When multiple chemical parameters are being monitored, it may be that the most spatially redundant well for one parameter is not the most redundant for others. The goal is then to remove only those wells that exhibit spatial redundancy *across* the monitored parameters. One possible strategy is to compute a separate set of global interpolation weights for each parameter and then average these sets across the parameters. That way, the final ranking for selecting candidate wells for removal will account for all the parameters of interest.

The strategy used for this investigation was to simply compare the lists of tentatively removed wells for each parameter and only remove those that appeared on each list. Remarkably, even though different spatial covariance models were fit to EDB and benzene for FS-12, the lists of spatially redundant wells were very similar for the two parameters at each of the weight thresholds tested. The close correspondence offers additional confidence that the wells targeted as candidates for removal did indeed provide redundant spatial information.

At Eastern Briarwood six distinct thresholds were tested with TCE and PCE. Again it was the case that despite the use of different spatial covariance models for the two chemicals, the lists of potentially redundant wells were very similar, exhibiting a high degree of overlap. Once again, the wells ultimately tagged for potential removal included only those flagged as redundant for *both* TCE and PCE.

**5) Assess Relative Uncertainty**

Although the weight thresholds provide an impartial way to identify potentially spatially redundant wells, the thresholds are arbitrary. Ultimately it must be determined whether removing those wells has

any measurable impact on the estimation of the indicator plume maps. One useful measure that is part of any standard kriging output is the local kriging variance, which, in the case of block kriging, is a separate number associated with each estimated local block. As with any statistical variance estimate, large local kriging variances suggest that the estimated value for a given block is much less precise than estimated blocks with small local kriging variances.

Since the size of the local kriging variance for a given block depends on the spatial covariance model, the number and configuration of the sampled locations, and the position of the estimated block relative to nearby samples (i.e., known well locations), the local kriging variance also provides a measure of relative spatial redundancy. In fact, by averaging the local kriging variances *across blocks* as suggested by Bertolino, et al (1983), the overall uncertainty using one configuration of well locations can be compared to the uncertainty derived from alternate configurations. This quantity — after further averaging across all the time slices — is denoted as the *global kriging variance* in the proposed optimization scheme.

In particular, if all wells with global kriging weights smaller than a particular threshold are eliminated from the mix, and the site is re-kriged on the same blocks, the new global kriging variance can be checked against the original measure to examine whether or not too much spatial information has been lost. Used in an iterative fashion, this algorithm allows the set of well locations to be narrowed to those that are most helpful as statistical estimators of ground-water quality. For instance, it might be agreed that a subset of wells can be removed from monitoring as long as the increase in global kriging variance is no more than, say, 5% of the original value. While a 5% increase may seem at first glance to be less than consequential, remember that the local kriging variances are being averaged across blocks and across time slices. Any increase in the overall average will necessarily entail a number of blocks with significant jumps in the local kriging variance, indicative of a loss of spatial information. So even small increases in the global kriging variance are likely to have significance.

Values of the global kriging variance for each weight threshold and its relative change with respect to the initial global kriging variance were determined and analyzed. Unfortunately, despite the simplicity of the global kriging variance, it must usually be supplemented by other uncertainty measures. In particular, it is quite helpful to compute ratios — on a block-by-block basis — of the local kriging variance at a given threshold and the initial local kriging variance before any well locations have been removed. By further averaging these ratios across time slices, it is possible to create maps of the time-averaged local kriging variance ratios.

Maps of the local kriging variance ratios indicate what parts of the estimated plume map are associated with the largest change in relative uncertainty, after removing a subset of the well locations. Presumably, by removing a subset of wells from the analysis, the kriging variances will tend to increase at some of the estimated blocks. Large ratios in particular sectors of the plume will then suggest that a significant amount of spatial information has been lost in those sectors. This allows further refining of the criteria used to remove wells.

As a final check of uncertainty, note that the final goal of the optimization is to ensure that reasonably consistent plume maps can be constructed even after removing a set of spatially redundant wells. Any weight threshold or targeted increase in the kriging variance is to some extent arbitrary. One

should therefore examine before-and-after indicator plume maps to determine if the basic pattern and features of the map estimates have been fundamentally altered. If so, it suggests that too many wells may have been removed from the mix.

When examining plume maps derived from this analysis, even subtle changes in contour lines indicate some loss of spatial information. However, examining such maps takes some practice and can be more subjective in nature than the checks on the local and global kriging variances outlined above.

**6) Finalize List of Redundant Wells**

After determining an appropriate threshold for each contaminant of concern, the lists of potentially redundant wells for each plume were compared. As noted earlier, there was a high degree of overlap in the lists comparing contaminants within each plume. Ultimately, only those well locations tagged on the lists of both contaminants at each plume were judged to be spatially redundant.

On the basis of the spatial analysis, 38 of 173 distinct locations were tagged as spatially redundant at FS-12. Of these 38, only 21 locations were still being monitored under the current regime as of Fall 1999. At Eastern Briarwood, 71 of 363 well locations were tagged as spatially redundant. None of these 71 are still being monitored according to information supplied by the latest contractor.

Despite the fact that many of the redundant well locations are no longer being regularly sampled, removing those that still exist and applying the recommended global sampling frequency (from the composite temporal variograms in Section 2) leads to significant potential cost reductions at FS-12 and Eastern Briarwood. The sampling and analysis budget at FS-12 could potentially be reduced by 42% from the current expenditure for a savings of approximately $165,000 per year. At Eastern Briarwood, the potential reduction is 36% for an annual savings of around $76,000.

It must be noted that the lists of spatially redundant wells were proposed for removal *strictly on the basis* of the above statistical analysis. Before such a recommendation is implemented, the specific well locations would need to be examined by hydrogeologists familiar with the sites and by the appropriate regulators to ensure that valuable information other than the concentration data used here would not be lost. Other than a change in cost estimates, the optimization algorithm would in no way be harmed or altered if someone decides for other reasons that one or more wells tagged as redundant should be kept on the monitoring list and not removed.

## Final Considerations

The benefits of optimization and any cost savings realized are strongly dependent on the quality and currency of the input data. As plumes change over time and historical data becomes regarded as unrepresentative of current conditions, it is highly recommended that an ongoing review be conducted, say, every three to five years after the initial implementation. At these intervals, the optimization algorithm should be re-conducted using more recent sampling information, in order to determine whether the global sampling frequency or the frequencies at individual wells need adjustment, and to

determine whether or not additional wells show significant spatial redundancies. It also might happen that new wells may need to be placed *into* the monitoring network.

To improve the success of ongoing reviews, operational adjustments to the sampling schedule should be entertained. By way of example, suppose the initial analysis suggests a global sampling interval of one year, so that each distinct well is to be sampled once per annum. Rather than sampling all wells at the same time every year, a better strategy would be to divide the wells into four non-overlapping subsets and sample one-quarter of the wells (i.e., one subset) each quarter. Such a scheme will tend to minimize biases or artifacts creeping into the data due to seasonal fluctuations, for instance.

In addition, so that enough pairs of measurements at different time lags are available to reconstruct a composite temporal variogram at the next program review, it is recommended that a random rotation be used to determine which subset of wells gets sampled during a given quarterly sampling event. In other words, if subset #1 was sampled during the third quarter of the first year after implementing the optimization scheme, it might be sampled during the first quarter of the following year, and perhaps the fourth quarter of the year after that. The intervals between consecutive samplings of the same well will then not always be a full year (sometimes less, sometimes more), but the sampling frequency will still be yearly.

By allowing for partially randomly-determined sampling intervals, data pairs can be formed at a variety of different lag times, thus enabling re-examination of the temporal variogram and whether or not the optimal sampling interval has changed. Otherwise, if a given well was sampled at precisely the same time from year to year, only pairs with a one-year lag time or greater could be formed.

On the spatial side of the algorithm, one way to determine whether new wells should be *added* to the network is to examine maps of the local kriging variances. Specific areas of the site with very high kriging variances represent parts of the plume where concentration estimates are likely to be rather uncertain. Often the placement of one or two wells in such areas will dramatically reduce the local kriging variances and improve the reliability of interpolated concentration maps made of the plume.

## References

Bertolino, F., Luciano, A., & Racugno, W. (1983) Some aspects of detection networks optimization with the kriging procedure. *Metron*, 41(3): 91-107.

Gilbert, R.O. (1987) *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.

Isaaks, E.H. & Srivastava, R.M. (1989) *Applied Geostatistics*. Oxford University Press, New York.

Johnson, V.M., Tuckfield, R.C., Ridley, M.N., & Anderson, R.A. (1996) Reducing the sampling frequency of ground-water monitoring wells. *Environmental Science & Technology*, 30(1): 355-358.

Tuckfield, R.C. (1994) Estimating an appropriate sampling frequency for monitoring ground water well contamination. Paper presented at International Nuclear Materials Management (INMM) Annual Meeting, Naples.